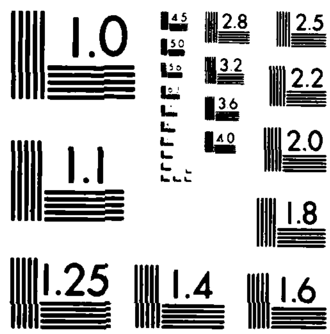
AD-A148 752 SMALL-SAMPLE PROPERTIES OF KERNEL DENSITY ESTIMATORS FROM RANDOMLY RIGHT-..(U) SOUTH CAROLINA UNIV COLUMBIA DEPT OF MATHEMATICS AND STATISTI.. W J PADGETT ET AL. OCT 84 TR-102 AFOSR-TR-84-1153 1/1

UNCLASSIFIED F/G 12/1 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Department of Mathematics
and Statistics
The University of South Carolina
Columbia, South Carolina 29208

SMALL-SAMPLE PROPERTIES OF KERNEL DENSITY ESTIMATORS FROM RANDOMLY RIGHT-CENSORED DATA

by

W. J. Padgett[1] and D. T. McNichols[2]

University of South Carolina

Statistics Technical Report No. 102
62G05-11

5

SMALL-SAMPLE PROPERTIES OF KERNEL DENSITY ESTIMATORS FROM RANDOMLY RIGHT-CENSORED DATA

by

W. J. Padgett[1] and D. T. McNichols[2]

University of South Carolina
Statistics Technical Report No. 102
62G05-11

October, 1984

[1] Department of Mathematics and Statistics, University of South Carolina, Columbia, SC 29208. Supported during the revision of this paper by the U. S. Air Force Office of Scientific Research and Army Research Office under grant number AFOSR-84-0156.

[2] Department of Statistics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Statistical TR No. 102	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 84-1153
6a. NAME OF PERFORMING ORGANIZATION University of South Carolina / 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research
6c. ADDRESS (City, State and ZIP Code) Department of Mathematics & Statistics Columbia SC 29208	7b. ADDRESS (City, State and ZIP Code) Directorate of Mathematical & Information Sciences, Bolling AFB DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR / 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-84-0156
8c. ADDRESS (City, State and ZIP Code) Bolling AFB DC 20332-6448	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. 61102F; PROJECT NO. 2304; TASK NO. A5; WORK UNIT NO.

11. TITLE (Include Security Classification)
SMALL-SAMPLE PROPERTIES OF KERNEL DENSITY ESTIMATORS FROM RANDOMLY RIGHT-CENSORED DATA

12. PERSONAL AUTHOR(S)
W.J. Padgett and D.T. McNichols

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Technical	FROM ______ TO ______	29 OCT 84	30

16. SUPPLEMENTARY NOTATION

17. COSATI CODES		
FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Random right-censorship; Monte Carlo simulations; Kernel estimation; Nonparametric density estimation; Bandwidth sequence.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The small-sample behavior of two kernel-type density estimators which have been proposed in the literature for randomly right-censored samples is investigated via Monte Carlo simulations. The extensive simulation study was performed for five families of life distributions, two different censoring distributions, three kernel functions, and several bandwidth sequences and for sample sizes from n=20 to n=300. The simulation results reinforce previous theoretical results for the estimators and lead to conjectures about their general behavior asymptotically as well as for small samples. A comparison of the two density estimators is also indicated.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL MAJ Brian W. Woodruff	22b. TELEPHONE NUMBER (Include Area Code) (202) 767- 5027 / 22c. OFFICE SYMBOL NM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

84 12 13 022

SMALL-SAMPLE PROPERTIES OF KERNEL DENSITY

ESTIMATORS FROM RANDOMLY RIGHT-CENSORED DATA

by

W. J. Padgett
Department of Mathematics
and Statistics
University of South Carolina
Columbia, S. C. 29208

D. T. McNichols
Department of Statistics
Virginia Polytechnic Institute
and State University
Blacksburg, Virginia 24061

ABSTRACT

The small-sample behavior of two kernel-type density estimators which have been proposed in the literature for randomly right-censored samples is investigated via Monte Carlo simulations. The extensive simulation study was performed for five families of life distributions, two different censoring distributions, three kernel functions, and several bandwidth sequences and for sample sizes from n=20 to n=300. The simulation results reinforce previous theoretical results for the estimators and lead to conjectures about their general behavior asymptotically as well as for small samples. A comparison of the two density estimators is also indicated.

1. INTRODUCTION

Density estimation is a very important topic in applied, as well as theoretical, statistics. In particular, nonparametric procedures for estimating an unknown density are extremely useful in determining the characteristics of a statistical population being sampled and have direct applications in many inference problems. The modern methods of nonparametric density estimation have been developed since the early 1950's and lead to smooth estimates which are more suitable for inference than simple histogram estimates. Most of these estimators were based on complete samples, that is, random samples of size n from the unknown density. There have been several reviews written which give extensive bibliographies of results on nonparametric density estimation from complete samples. For example, see Wegman (1972 a,b), Fryer (1977), Tapia and Thompson (1978), Wertz and Schneider (1979), and Bean and Tsokos (1980).

Recently, density estimation from incomplete or censored samples has received a great deal of attention. Right-censored observations arise in many life testing situations and are very common in survival analysis (Lagakos, 1979). Such data occur often in medical trials when patients may enter treatment at different times and then either die from the disease, or cause, under investigation or leave the study before it is terminated (move away or die from another competing cause). Also, in industrial life testing, items may be removed from the study at various times for more extensive analysis or for other reasons. For such situations, it is of interest to obtain nonparametric estimates of the density function of the lifetime variable based on the right-censored data. The development of such density (or related function) estimates has only recently been considered, and a survey of known results was given by Padgett and McNichols (1984). The developments for censored data have followed the same basic approaches as for the complete-sample case but generally present greater mathematical difficulties.

Kernel density estimators from randomly right-censored data have been studied by several authors. A kernel-type density estimator was proposed by Blum and Susarla (1980) and its asymptotic properties were studied. In particular, the asymptotic theory of the maximum deviation of their estimator was presented, extending the results of Rosenblatt (1976) to the case of random right-censorship. The strong consistency properties of the kernel estimator based on the product-limit estimate of the distribution function were studied by Földes, Rejtö and Winter (1981). McNichols and Padgett (1981) obtained very complicated finite-sample expressions for the kernel density estimator and showed that it was asymptotically unbiased and that its variance approached zero as the sample size increased, assuming the Koziol and Green (1976) model of random censorship. Also, a modification of the kernel density estimator in which the bandwidth depended on the data was proposed by McNichols and Padgett (1984). However, only asymptotic properties were obtained in all of these results, except for those by McNichols and Padgett (1981) with respect to the Koziol-Green model which is somewhat restrictive in practice.

It is the purpose of this paper to study, by fairly extensive Monte Carlo simulations, the finite-sample behavior of kernel density estimators based on randomly right-censored data. The simulation study was performed since it is very difficult, if not impossible, to obtain (even approximate) expressions for the biases, mean-squared errors, variances, and sampling distributions of such estimators for finite sample sizes under general nonrestrictive conditions. Several different families of lifetime distributions, various types of censoring distributions that are assumed in practice, various bandwidth sequences, and three different kernel functions were used in the simulations. Since, for censored data, optimal bandwidth results analogous to those for complete samples is not available, some attention is given in the simulations to the behavior of the estimators with respect to the bandwidth.

Randomly right-censored data and the product-limit estimator will be discussed in Section 2. The kernel density estimators that are to be studied will be given in Section 3. The computer simulations will be described and a representative proportion of the simulation results will be given in Section 4. Finally, in Section 5 some conclusions concerning the small-sample behavior of the kernel estimators studied will be stated or conjectured.

2. RANDOMLY RIGHT-CENSORED SAMPLES

Let $X_1^o, X_2^o, \ldots, X_n^o$ denote the true survival times of n items or individuals which are censored on the right by a sequence $U_1, U_2, \ldots, U_n$ which in general may be either constants or random variables. It is assumed that the X_i^o's are nonnegative independent identically distributed random variables with common unknown distribution function F^o. For the problem of density estimation, it is assumed that F^o is absolutely continuous with density f^o.

The observed right-censored data are denoted by the pairs (X_i, Δ_i), $i=1,\ldots,n$, where

$$X_i = \min\{X_i^o, U_i\}, \quad \Delta_i = \begin{cases} 1 & \text{if } X_i^o \leq U_i \\ 0 & \text{if } X_i^o > U_i . \end{cases}$$

Thus, it is known which observations are times of failure or death and which ones are censored or loss times. The nature of the censoring mechanism depends on the U_i's: (i) If $U_1, \ldots, U_n$ are fixed constants, the observations are time-truncated. If all U_i's are equal to the same constant, then the case of Type I censoring results. (ii) If all $U_i = X^o_{(r)}$, the $r\underline{th}$ order statistic of $X^o_1,\ldots,X^o_n$, then the situation is that of simple Type II censoring. (iii) If $U_1,\ldots,U_n$ constitute a random sample from a distribution H (which is usually unknown) and are independent of $X^o_1,\ldots,X^o_n$, then (X_i,Δ_i), $i=1,2,\ldots,n$, is called a <u>randomly</u> <u>right-censored</u> <u>sample</u>.

The random censorship model (iii) is assumed for the results presented here. It is attractive because of its mathematical convenience. Assuming this model, $\Delta_1,\ldots,\Delta_n$ are independent Bernoulli random variables and the distribution function F of each X_i, $i=1,\ldots,n$, is given by $1-F = (1-F^o)(1-H)$. Under the Koziol and Green (1976) model of random censorship, which is the proportional hazards assumption of Cox (1972), it is assumed that there is a positive constant β such that $1-H = (1-F^o)^\beta$. Then by a result of Chen, Hollander, and Langberg (1982), the pairs (X^o_i,U_i), $i=1,\ldots,n$, follow the proportional hazards model if and only if $(X_1,\ldots,X_n)$ and $(\Delta_1,\ldots,\Delta_n)$ are independent. This Koziol-Green model of random censorship arises in several situations (Efron, 1967; Csörgo and Horváth, 1981; Chen, Hollander and Langberg, 1982). Note that β is a censoring coefficient since $a = P(X^o_i \leq U_i) = (1 + \beta)^{-1}$, which is the probability of an uncensored observation.

Based on the censored sample (X_i,Δ_i), $i=1,\ldots,n$, a popular estimator of the survival probability $S^o(t) = 1-F^o(t)$ at $t \geq 0$ is the product-limit estimator, proposed by Kaplan and Meier (1958) as the "nonparametric maximum likelihood estimator" of S^o. This estimator was shown to be "self-consistent" by Efron (1967).

Let (Z_i,Δ'_i), $i=1,\ldots,n$, denote the ordered X_i's along with their corresponding Δ_i's. A value of the censored sample will be denoted by the corresponding lower case letters (x_i,δ_i) or (z_i,δ_i') for the unordered or ordered sample, respectively. The product-limit estimator of S^o is defined by (Efron, 1967)

$$\hat{P}_n(t) = \begin{cases} 1, & 0 \leqq t \leqq Z_1 \\ \prod_{i=1}^{k-1} \left(\frac{n-i}{n-i+1}\right)^{\delta_i'}, & t \quad (Z_{k-1}, Z_k],\ k=2,\ldots,n. \\ 0, & t > Z_n. \end{cases}$$

Denote the product-limit estimator of $F^o(t)$ by $\hat{F}_n(t) = 1-\hat{P}_n(t)$, and let s_j denote the jump of $\hat{P}_n$(or $\hat{F}_n$) at Z_j, that is,

$$s_j = \begin{cases} 1-\hat{P}_n(Z_2), & j=1 \\ \hat{P}_n(Z_j) - \hat{P}_n(Z_{j+1}), & j=2,\ldots,n-1 \\ \hat{P}_n(Z_n), & j=n. \end{cases}$$

Note that $s_j=0$ if and only if $\delta_j'=0$, $j < n$, that is, if Z_j is a censored observation.

The product-limit estimator has played a central role in the analysis of censored survival data (Miller, 1981), and its properties have been studied extensively by many authors, for example, Breslow and Crowley (1974), Földes, Rejtö and Winter (1980), and Wellner (1982).

3. THE KERNEL DENSITY ESTIMATORS

Since the work of Rosenblatt (1956) and Parzen (1962), kernel density estimators have been perhaps the most popular density estimators used in practice and have been studied extensively regarding their theoretical properties. Also, various modifications with respect to the bandwidth sequence and kernel have been proposed. Until recently, all of the results were for complete samples (see Fryer, 1977, or Bean and Tsokos, 1980). For randomly right-censored data, the first results for kernel density estimation did not appear until 1980.

Blum and Susarla (1980) generalized the complete-sample results of Rosenblatt (1976) concerning maximum deviation of density estimates by the kernel method. They obtained limit theorems for the maximum over a finite interval of a normalized deviation of the density estimate when the observations were censored on the right. The results were useful for goodness-of-fit tests and tests of hypothesis about the unknown lifetime density f^o. To define the Blum-Susarla estimator based on the randomly censored observations (X_i, Δ_i), $i=1,\ldots,n$, let $\{h=h(n)\}$ be a positive sequence converging to zero as $n \to \infty$ and let

$$N^+(x) = \text{number of } X_i\text{'s} > x .$$

Define

$$H_n^*(x) = \prod_{j=1}^{n} \left\{ \frac{1+N^+(X_j)}{2+N^+(X_j)} \right\}^{I[\delta_j=0, X_j \leq x]} ,$$

where $I_{[A]}$ denotes the indicator function of the measurable set A. By a modification of the product-limit estimator, it can be shown that H_n^* is a good estimate of the survival function for the censoring distribution, $H^* = 1-H$ (Blum and Susarla, 1980). For a kernel function K satisfying certain conditions, the Blum-Susarla estimator is defined by

$$f_n^*(x) = \frac{1}{nh} \cdot \frac{\sum_{j=1}^{n} K(\frac{x-X_j}{h}) \; I_{[\delta_j=1]}}{H_n^*(x)}. \tag{3.1}$$

By following standard arguments, $(f^o H^*)_n(x) = (nh)^{-1} \Sigma_{j=1}^{n} K((x-X_j)/h) I_{[\delta_j=1]}$ and $H_n^*(x)$ can be shown to be good estimators of $f^o(x)H^*(x)$ and $H^*(x)$, respectively. This motivates the use of (3.1) as an estimator of $f^o(x)$.

The main results of Blum and Susarla (1980) concern the asymptotic distribution of

$$M_n = (nh)^{\frac{1}{2}} \sup_{0 \leq x \leq 1} \frac{|f_n^*(x) - [hH^*(x)]^{-1} E[K(\frac{x-X_1}{h}) I_{[\delta_1=1]}]|}{[f^o(x)/H^*(x)]^{\frac{1}{2}}}$$

under various conditions on f^o, K, and H.

Földes, Rejtö and Winter (1981) obtained strong convergence results for the kernel density estimator

$$\hat{f}_n(x) = h^{-1} \int_{-\infty}^{\infty} K(\frac{x-t}{h})\, d\hat{F}_n(t), \tag{3.2}$$

which reduces to the usual Parzen (1962) density estimator in the case of no censoring (since the product-limit estimator $\hat{F}_n$ reduces to the usual empirical distribution function). Their results were obtained under various conditions on H, F^o, f^o, and K, and they assumed that the bandwidth sequence $\{h(n)\}$ was such that $h(n) \to 0$ but $h(n)(n/\log(n))^{1/8} \to \infty$ as $n \to \infty$.

McNichols and Padgett (1981) wrote (3.2) as

$$\hat{f}_n(x) = h^{-1} \sum_{j=1}^{n} s_j\, K(\frac{x-Z_j}{h}), \tag{3.3}$$

where Z_j is the $j\underline{th}$ ordered observation and s_j denotes the jump of $\hat{F}_n$ at Z_j. They considered the mean, variance, and mean-squared error of (3.3) under the Koziol-Green model. Expressions for the mean and variance of (3.3) at each $x \geqq 0$ were obtained and asymptotic unbiasedness and mean-square convergence was shown with K and $\{h(n)\}$ satisfying the usual Parzen (1962) conditions. Note that the sums in both (3.1) and (3.3) only explicitly include the terms with uncensored observations although the censoring is treated somewhat differently.

The small-sample properties of (3.1) and (3.3) have not been studied previously, either analytically or by computer simulations, other than under the restrictions of the Koziol-Green model (McNichols and Padgett, 1981). In the next section, a rather extensive Monte Carlo simulation study of the estimators (3.1) and (3.3) for small sample sizes will be described, and some representative results will be presented.

It should be mentioned that a modification of $\hat{f}_n$ in which the bandwidth h is data-driven has been given by McNichols and Padgett (1984). It was shown that if $h = h(X_1,\ldots,X_n)$ is a "nearest neighbor" type function, then the conditions for consistency of the modified estimator hold. Also, it should be remarked that the data-based algorithms for choosing h in the complete sample case discussed by Scott and Factor (1981) do not seem to be fruitful for the case of censored samples. In particular, an expression similar to their (2.4) (see also Parzen, 1962), and hence (2.10), is not available in the censored data case and seems to be extremely difficult to obtain (McNichols and Padgett, 1981). A likelihood approach corresponding to their expression (2.8) does not seem to be feasible either, since for censored data, the survival function corresponding to $\hat{f}_n$ appears in the likelihood function. Hence, in the simulation study described in the next section, some attention is given to estimating the mean squared errors of the kernel estimators as a function of various bandwidth values. This gives an indication of the range of values of h which tend to minimize mean squared errors of both (3.1) and (3.3) in the cases simulated.

4. THE MONTE CARLO SIMULATIONS

Simulations were performed for randomly right-censored samples generated from five different families of life distributions commonly used in the literature: exponential with mean β, denoted $E(\beta)$, gamma with parameters α and β, denoted $G(\alpha,\beta)$, Weibull with density $f(x;\ \alpha,\ \beta) = \frac{\alpha}{\beta}(\frac{x}{\beta})^{\alpha-1} \exp[-(x/\beta)^{\alpha}]$, $x > 0$, denoted $W(\alpha,\beta)$, lognormal with mean $\exp(\alpha + \frac{1}{2}\beta^2)$, denoted $L(\alpha,\beta)$, and inverse Gaussian with density $f(x;\ \mu,\ \lambda) = [\lambda/(2\pi x^3)]^{\frac{1}{2}} \exp[-\lambda(x-\mu)^2/(2\mu^2 x)]$, $x > 0$, denoted $IG(\mu,\lambda)$. Two different types of censoring distributions were utilized, exponential with mean one and uniform on $(0,t_q)$, where t_q denotes the q^{th} percentile of the life distribution, $0 < q < 100$. Three different kernel functions K were used, the standard normal density, the uniform density on $[-1,1]$, and the triangular density on $[-1,1]$,

$$K(x) = \begin{cases} 1 - |x|, & |x| \leqq 1 \\ 0, & \text{otherwise.} \end{cases}$$

In addition, several bandwidth values $h = h(n)$ were used in the study, including $h(n) = n^{-p}$ for various values of p.

The simulations represented in Tables 4.1-4.7 and 4.12-4.15 were based on 1,000 randomly right-censored samples each of size n for each choice of life distribution, censoring distribution, kernel function, and bandwidth value for $n = 20, 50, 100$, and 300. For each sample, the estimates (3.1) and (3.3) were computed for values of t = 5th, 10th, 20th,...,90th, and 95th percentiles of the censoring distribution (t values of 5th, 50th, and 95th percentiles only are reported in these tables). At each t, the bias, mean squared error (MSE), and variance of the estimators were estimated from the 1,000 computed values. Also, the standard error of the estimate of MSE at each t was computed for each estimator. The standard errors were bounded by 10^{-2}.

The computer programs for the simulations were written in Fortran on an Amdahl 470 V611 computer. The random number generators contained in the

International Mathematical and Statistical Libraries (1980)(IMSL) package were used in the generation of the required samples. Uniform random numbers were generated from the IMSL subroutine GGUBS. IMSL subroutine GGEXN was used for the exponential random numbers, GGAMR for gamma, GGWIB for Weibull, and GGNLG for lognormal random numbers. To generate a value x from the inverse Gaussian distribution, the procedure given by Michael, Schucany, and Haas (1976) was used.

The Monte Carlo simulations were performed in the following manner: A random sample $X_1^o,\ldots,X_n^o$ was generated from the life distribution, and a random sample $U_1,\ldots,U_n$ was generated from the censoring distribution. Next, the randomly right-censored sample (X_i,Δ_i), $i=1,\ldots,n$, was obtained by $X_i = \min\{X_i^o,U_i\}$, $\Delta_i = 1$ if $X_i = X_i^o$ and $\Delta_i = 0$ if $X_i = U_i$. The values $X_1,\ldots,X_n$ were ordered to yield (Z_i,Δ_i'), $i=1,\ldots,n$, and the product-limit estimator was computed along with the jump size s_i at each Z_i. The estimators $f_n^*(t)$ and $\hat{f}_n(t)$ given by (3.1) and (3.3) were computed at the appropriate values of t. This entire procedure was repeated for 1,000 randomly right-censored samples. The average biases, mean squared errors, and variances as well as the standard errors (all were bounded by 10^{-2}) of the estimated mean squared errors were computed for $f_n^*(t)$ and $\hat{f}_n(t)$ over the 1,000 samples. The entire procedure was repeated for each sample size, life distribution, censoring distribution, kernel function, and bandwidth value mentioned before.

Some representative simulation results for $\hat{f}_n(t)$ are given in Tables 4.1 - 4.7. All of the results cannot be listed due to space limitations. All entries in all tables are to be multiplied by 10^{-4}.

In the hope of gaining some insight into the behavior of f_n^* and $\hat{f}_n$ with respect to the bandwidth values h, several cases were simulated, using 200 samples each (instead of 1000, due to computer time constraints), in which the estimated MSE was obtained as a function of h. The estimated MSE was obtained for f_n^* and $\hat{f}_n$ at values $h = .05,(.05),.55$. For sample size $n=50$ and 100 some representative results are shown in Tables 4.8-4.11. Note that the range of h values contains $n^{-1/2}$, $n^{-1/3}$, and $n^{-1/5}$ within the boundaries for these sample sizes. The results indicate that f_n^* and $\hat{f}_n$ tend to behave similarly with respect to MSE. Therefore, in order to indicate a comparison of the behavior of f_n^* and $\hat{f}_n$ as density estimators when the same bandwidth values are used, some representative simulation results are listed in Tables 4.12-4.15. In these tables a = P(an uncensored observation) $= P(X_i^o \leq U_i)$.

5. CONCLUSIONS

Several conclusions concerning the small-sample behavior of the kernel density estimators $\hat{f}_n$ and f_n^* can be stated based on the extensive simulations described in Section 4. In particular, the simulation results indicate the following for $\hat{f}_n$: The estimated variances of $\hat{f}_n(t)$ increase as the bandwidth sequence $\{h(n)\}$ varies from $n^{-1/5}$ to $n^{-1/2}$. For $h(n) = n^{-1/5}$, the variances of $\hat{f}_n(t)$ decrease as n increases, but probably do not converge to zero uniformly in t. For small values of t, the bias of $\hat{f}_n(t)$ is larger in magnitude than the biases for moderate to large values of t. Overall, with respect to the criterion of mean squared error, for both f_n^* and $\hat{f}_n$ with moderate to large t, $h(n) = n^{-1/5}$ appears to be the best choice for the bandwidth among the values $h(n) = n^{-p}$, $p = 1/2, 1/3, 1/5$, whereas for small t, $n^{-1/2}$ or $n^{-1/3}$ appears better with respect to mean squared error (Tables 4.1-4.7). This is supported by the representative results in Tables 4.8-4.11. Of the three kernel functions studied, the

Table 4.1 Density Estimate $\hat{f}_n$.

Life Distribution: E(1), Censoring Distribution: E(1).
(All entries are to be multiplied by 1.0E-04.)

$f^o(t)$ (t)	n	p	Standard Normal Kernel Bias	Var. $\hat{f}_n(t)$	MSE. $\hat{f}_n(t)$	Triangular Kernel Bias	Var. $\hat{f}_n(t)$	MSE. $\hat{f}_n(t)$	Uniform Kernel Bias	Var. $\hat{f}_n(t)$	MSE. $\hat{f}_n(t)$
9500 (513)	20	1/5	- 5852	63	3488	- 4600	255	2371	- 5344	156	3012
		1/3	- 5286	121	2915	- 3979	473	2056	- 4808	276	2588
		1/2	- 4564	268	2351	- 3051	984	1914	- 4104	548	2232
	100	1/5	- 5457	21	2998	- 4149	87	1809	- 4967	46	2513
		1/3	- 4553	58	2132	- 3041	217	1142	- 4066	119	1773
		1/2	- 3206	192	1220	- 1227	608	759	- 2501	361	986
	300	1/5	- 5125	10	2636	- 3759	37	1450	- 4635	22	2171
		1/3	- 3940	32	1584	- 2195	110	592	- 3408	65	1226
		1/2	- 1899	126	487	- 37	351	351	- 551	236	266
5000 (6931)	20	1/5	382	84	98	273	413	420	391	262	277
		1/3	172	189	192	190	731	734	250	515	520
		1/2	263	438	445	160	1366	1368	159	967	969
	100	1/5	- 93	29	30	25	116	116	100	77	78
		1/3	77	81	82	- 31	260	260	0	173	172
		1/2	- 20	232	232	- 76	626	626	- 13	456	455
	300	1/5	86	14	14	25	52	52	69	37	37
		1/3	39	46	46	16	129	129	6	94	94
		1/2	16	143	143	- 13	365	365	13	261	261
500 (29957)	20	1/5	- 117	34	36	- 265	71	78	- 232	65	71
		1/3	- 214	49	54	- 291	90	98	- 255	94	101
		1/2	- 267	72	80	- 348	106	118	- 283	145	153
	100	1/5	44	31	31	- 42	94	94	- 20	67	67
		1/3	- 22	69	69	- 46	193	193	- 44	131	131
		1/2	- 48	175	175	- 39	473	473	- 77	295	295
	300	1/5	102	19	20	88	68	69	90	46	47
		1/3	90	60	61	99	166	166	99	124	125
		1/2	99	181	182	88	444	445	85	327	328

Table 4.2 Density Estimate $\hat{f}_n$.

Life Distribution: E(1), Censoring Distribution: $U(0, t_{90})$, $t_{90} = 2.3026$.
(All entries are to be multiplied by 1.0E-04.)

$f^o(t)$ (t)	n	p	Standard Normal Kernel Bias	Var.$\hat{f}_n(t)$	MSE.$\hat{f}_n(t)$	Triangular Kernel Bias	Var.$\hat{f}_n(t)$	MSE.$\hat{f}_n(t)$	Uniform Kernel Bias	Var.$\hat{f}_n(t)$	MSE.$\hat{f}_n(t)$
8913 (1151)	20	1/5	- 5076	53	2629	- 3338	249	1364	- 4456	128	2114
		1/3	- 4316	112	1975	- 2373	476	1039	- 3678	231	1584
		1/2	- 3227	271	1313	- 1021	987	1091	- 2361	582	1139
	100	1/5	- 4500	18	2043	- 2639	80	777	- 3849	42	1524
		1/3	- 3191	55	1073	- 1052	202	313	- 2355	109	663
		1/2	- 1233	181	333	24	532	532	82	397	397
	300	1/5	- 4057	9	1656	- 2092	40	478	- 3385	21	1167
		1/3	- 2284	35	557	- 223	115	119	- 1107	70	192
		1/2	- 197	130	134	43	347	347	24	249	249
3162 (11513)	20	1/5	573	44	77	197	283	286	398	204	219
		1/3	427	128	146	55	499	499	207	358	362
		1/2	177	299	302	- 38	922	922	24	690	690
	100	1/5	295	22	31	68	79	80	103	54	55
		1/3	94	57	58	54	172	172	54	121	121
		1/2	53	154	154	21	399	399	58	290	290
	300	1/5	153	9	11	3	32	32	37	24	24
		1/3	12	28	28	- 26	83	83	- 24	58	58
		1/2	- 27	92	92	- 32	242	242	- 30	180	180
1122 (21874)	20	1/5	465	64	86	632	374	414	409	176	193
		1/3	501	156	181	821	729	796	593	375	409
		1/2	659	406	449	1106	1515	1636	826	877	944
	100	1/5	74	27	82	1703	206	496	946	61	151
		1/3	1380	123	313	2714	739	1475	1874	295	646
		1/2	2600	646	1322	3423	2786	3955	3447	1698	2884
	300	1/5	975	12	107	2235	76	575	1461	25	238
		1/3	2134	63	519	3472	453	1659	3197	155	1177
		1/2	3316	667	1766	2604	3461	4136	2816	2485	3277

Table 4.3 Density Estimate $\hat{f}_n$.

Life Distribution: G(2,1), Censoring Distribution: E(1).
(All entries are to be multiplied by 1.0E-04.)

$f^o(t)$ (t)	n	p	Standard Normal Kernel Bias	Var.$\hat{f}_n(t)$	MSE.$\hat{f}_n(t)$	Triangular Kernel Bias	Var.$\hat{f}_n(t)$	MSE.$\hat{f}_n(t)$	Uniform Kernal Bias	Var.$\hat{f}_n(t)$	MSE.$\hat{f}_n(t)$
487 (513)	20	1/5	802	36	101	417	53	71	639	68	109
		1/3	624	42	81	263	63	70	438	75	94
		1/2	401	52	68	123	88	89	229	87	92
	100	1/5	650	8	50	280	12	20	455	14	35
		1/3	374	10	24	100	18	19	207	18	23
		1/2	123	16	18	- 3	33	33	19	27	27
	300	1/5	547	3	33	198	5	9	356	5	18
		1/3	232	5	10	28	8	8	95	7	8
		1/2	19	9	9	- 24	20	20	- 22	15	15
3466 (6931)	20	1/5	- 648	125	167	- 239	359	365	- 385	280	295
		1/3	- 437	208	227	- 131	587	588	- 242	404	410
		1/2	- 228	380	385	- 58	1005	1005	- 41	784	784
	100	1/5	- 501	31	56	- 115	91	92	- 217	66	70
		1/3	- 181	67	71	- 37	184	184	- 62	131	131
		1/2	- 46	165	165	- 22	429	429	4	322	322
	300	1/5	- 339	14	25	- 69	41	42	- 123	28	29
		1/3	- 84	37	37	- 18	95	95	- 34	68	68
		1/2	- 19	105	105	- 41	252	252	- 14	193	193
1498 (29957)	20	1/5	30	187	187	- 141	644	646	- 99	465	465
		1/3	- 81	362	363	- 202	1010	1013	- 210	720	724
		1/2	- 148	678	680	- 202	1651	1653	- 204	1313	1316
	100	1/5	451	129	149	512	596	621	481	356	379
		1/3	497	408	432	561	1382	1412	517	904	930
		1/2	544	1230	1259	637	3576	3614	2303	2327	242
	300	1/5	235	74	79	254	264	270	239	163	168
		1/3	252	228	235	322	742	752	259	462	468
		1/2	325	821	831	382	2319	2332	360	1488	1499

Table 4.4 Density Estimate $\hat{f}_n$.

Life Distribution: G(2,1), Censoring Distribution: $U(0, t_{90})$, $t_{90} = 3.8897$.
(All entries are to be multiplied by 1.0E-04.)

$f^o(t)$ (t)	n	p	Standard Normal Kernel Bias	Var.$\hat{f}_n(t)$	MSE.$\hat{f}_n(t)$	Triangular Kernel Bias	Var.$\hat{f}_n(t)$	MSE.$\hat{f}_n(t)$	Uniform Kernel Bias	Var.$\hat{f}_n(t)$	MSE.$\hat{f}_n(t)$
1601 (1944)	20	1/5	20	34	34	- 84	90	91	- 29	72	72
		1/3	- 31	55	55	- 99	135	136	- 110	104	104
		1/2	- 86	93	93	- 57	232	232	- 115	169	170
	100	1/5	- 16	9	9	- 85	26	27	- 92	18	19
		1/3	- 76	19	20	- 40	53	53	- 90	37	37
		1/2	- 47	48	48	- 13	119	119	- 8	88	88
	300	1/5	- 53	4	4	- 98	10	11	- 124	7	9
		1/3	- 96	9	10	- 32	25	25	- 64	18	18
		1/2	- 31	28	28	- 28	64	64	- 19	49	49
2781 (19448)	20	1/5	64	75	75	55	309	309	26	198	198
		1/3	59	154	155	67	509	509	65	355	355
		1/2	55	324	324	76	908	908	15	624	624
	100	1/5	5	22	22	36	75	75	33	52	52
		1/3	31	55	55	42	153	153	36	114	114
		1/2	42	138	138	61	367	367	71	267	268
	300	1/5	15	10	10	41	33	33	32	23	23
		1/3	41	29	29	36	81	81	47	56	56
		1/2	30	88	88	- 30	213	213	- 17	158	158
917 (36952)	20	1/5	309	87	97	455	404	424	312	213	222
		1/3	360	189	202	592	723	757	404	403	418
		1/2	481	436	459	736	1349	1402	657	896	939
	100	1/5	741	38	93	1557	239	481	1262	76	235
		1/3	1324	150	325	2069	759	1186	1902	381	742
		1/2	1993	668	1064	2087	2497	2930	2148	1733	2193
	300	1/5	971	17	111	1884	99	454	1404	38	235
		1/3	1795	87	409	1722	658	954	2110	472	917
		1/2	1667	752	1029	1295	2205	2371	1351	1724	1905

Table 4.5 Density Estimate $\hat{f}_n$.

Life Distribution: IG(3,1), Censoring Distribution: E(1).
(All entries are to be multiplied by 1.0E-04.)

$f^o(t)$ (t)	n	p	Standard Normal Kernel Bias	Var.$\hat{f}_n(t)$	MSE.$\hat{f}_n(t)$	Triangular Kernel Bias	Var.$\hat{f}_n(t)$	MSE.$\hat{f}_n(t)$	Uniform Kernel Bias	Var.$\hat{f}_n(t)$	MSE.$\hat{f}_n(t)$
28 (513)	20	1/5	2240	61	563	2102	121	563	2433	133	724
		1/3	2275	87	604	1724	138	435	2345	182	732
		1/2	2049	119	539	1105	139	261	1694	236	523
	100	1/5	2251	14	521	1781	24	342	2304	32	563
		1/3	1986	21	416	1056	25	137	1650	44	316
		1/2	1155	24	157	357	18	30	663	39	83
	300	1/5	2216	6	497	1537	9	245	2126	12	464
		1/3	1628	8	273	673	8	53	1139	15	145
		1/2	598	7	43	138	5	7	269	10	18
4513 (6931)	20	1/5	- 873	141	217	101	428	428	46	319	319
		1/3	- 301	243	252	25	689	689	174	506	508
		1/2	34	443	443	15	1280	1278	- 118	855	855
	100	1/5	- 557	34	65	55	115	115	77	74	74
		1/3	8	82	82	25	248	248	42	181	181
		1/2	28	222	221	36	558	558	- 7	417	416
	300	1/5	- 239	16	22	64	52	52	107	36	37
		1/3	65	46	46	29	126	126	1	89	89
		1/2	37	138	138	103	343	343	51	255	255
769 (29957)	20	1/5	426	159	177	339	491	502	407	384	400
		1/3	383	288	302	275	724	730	361	626	639
		1/2	329	506	517	174	1137	1139	268	987	994
	100	1/5	669	134	179	714	488	538	712	330	380
		1/3	709	352	402	707	1026	1075	720	747	798
		1/2	712	918	968	705	2456	2503	687	1744	1789
	300	1/5	189	65	68	117	172	173	151	127	129
		1/3	124	151	152	103	423	424	108	294	294
		1/2	104	462	463	184	1546	1548	108	759	759

Table 4.6 Density Estimate $\hat{f}_n$.

Life Distribution: W(0.5,1), Censoring Distribution: E(1).
(All entries are to be multiplied by 1.0E-04.)

$f^o(t)$ (t)	n	p	Standard Normal Kernel Bias	Var.$\hat{f}_n(t)$	MSE.$\hat{f}_n(t)$	Triangular Kernel Bias	Var.$\hat{f}_n(t)$	MSE.$\hat{f}_n(t)$	Uniform Kernel Bias	Var.$\hat{f}_n(t)$	MSE.$\hat{f}_n(t)$
17603 (513)	20	1/5	- 13494	76	18287	- 10182	349	10717	- 12644	147	16134
		1/3	- 12203	153	15046	- 7978	708	7073	- 11048	296	12504
		1/2	- 10004	376	10383	- 4470	1631	3628	- 8399	727	7781
	100	1/5	- 12548	23	15770	- 8500	117	7342	- 11455	47	13171
		1/3	- 9871	75	9819	- 4210	334	2106	- 8226	142	6910
		1/2	- 4714	299	2521	2313	1044	1579	- 1426	575	778
	300	1/5	- 11704	12	13710	- 7108	58	5111	- 10461	23	10966
		1/3	- 7685	49	5956	- 1078	203	319	- 5488	93	3106
		1/2	50	240	240	3631	681	1999	6905	533	5302
2612 (6931)	20	1/5	1124	89	215	400	297	313	812	248	313
		1/3	1045	160	269	185	469	473	373	371	384
		1/2	433	305	324	79	803	803	114	612	612
	100	1/5	879	22	100	85	72	73	243	54	60
		1/3	243	53	59	- 21	142	142	5	105	105
		1/2	- 8	128	128	- 43	327	327	- 47	247	247
	300	1/5	697	10	59	58	33	33	157	24	26
		1/3	89	29	30	- 25	72	72	2	56	56
		1/2	- 28	80	80	- 39	190	190	- 39	147	147
512 (29957)	20	1/5	225	83	88	204	263	267	197	186	190
		1/3	203	147	151	226	429	434	167	279	281
		1/2	206	275	279	263	776	783	245	536	542
	100	1/5	529	77	105	602	301	337	574	191	224
		1/3	585	211	245	640	681	721	581	440	474
		1/2	630	604	643	706	1758	1806	666	1163	1206
	300	1/5	268	53	60	245	159	165	264	114	121
		1/3	252	143	149	224	351	356	250	275	281
		1/2	211	377	381	124	783	783	162	630	632

Table 4.7 Density Estimate $\hat{f}_n$.

Life Distribution: W(2,1), Censoring Distribution: E(1).
(All entries are to be multiplied by 1.0E-04.)

$f^o(t)$ (t)	n	p	Standard Normal Kernel			Triangular Kernel			Uniform Kernel		
			Bias	Var. $\hat{f}_n(t)$	MSE. $\hat{f}_n(t)$	Bias	Var. $\hat{f}_n(t)$	MSE. $\hat{f}_n(t)$	Bias	Var. $\hat{f}_n(t)$	MSE. $\hat{f}_n(t)$
1023 (513)	20	1/5	2008	43	446	1122	116	242	1754	137	445
		1/3	1682	74	357	673	151	196	1193	177	320
		1/2	1071	114	229	257	198	204	552	219	249
	100	1/5	1730	13	312	762	28	86	1269	34	196
		1/3	1029	23	129	280	38	46	576	41	74
		1/2	342	35	47	7	70	70	107	63	64
	300	1/5	1504	5	231	559	9	40	989	11	109
		1/3	653	8	51	118	16	17	290	14	22
		1/2	94	17	18	- 34	39	40	- 21	31	31
8574 (6931)	20	1/5	- 2976	46	932	- 668	468	512	-1481	225	446
		1/3	- 1725	161	458	- 196	998	1001	- 636	545	585
		1/2	- 641	517	558	107	2148	2148	- 28	1049	1047
	100	1/5	- 2104	25	408	- 495	154	178	- 911	90	173
		1/3	- 798	100	164	- 187	387	390	- 330	257	268
		1/2	- 226	341	346	- 95	1016	1016	- 138	692	694
	300	1/5	- 1500	13	238	- 326	69	80	- 581	41	75
		1/3	- 407	58	75	- 147	210	212	- 199	131	135
		1/2	- 146	240	242	- 75	676	676	- 168	485	488
7 (29957)	20	1/5	61	1	2	- 3	0	0	- 1	1	1
		1/3	5	1	1	- 6	0	0	0	3	3
		1/2	- 4	0	0	- 7	0	0	- 5	0	0
	100	1/5	22	0	0	0	0	0	1	0	0
		1/3	1	0	0	0	1	1	- 1	1	1
		1/2	0	1	1	- 1	2	2	1	4	4
	300	1/5	9	0	0	- 3	0	0	- 2	0	0
		1/3	- 2	0	0	- 4	0	0	- 3	0	0
		1/2	- 4	0	0	- 3	0	0	- 3	1	1

Table 4.8. Estimated MSE of Kernel Density Estimators

Life Distribution: E(1), Censoring Distribution: $U(0,t_{.90})$
Kernel: N(0,1)
(All entries are to be multiplied by 1.0E-04.)

(a) n=50

p (t_p) \ h		.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55
.10	a.	830	390	240	310	420	610	760	980	1130	1280	1450
(.23)	b.	820	390	250	330	470	700	900	1160	1350	1530	1760
.25	a.	820	390	190	110	100	50	50	60	70	100	1400
(.58)	b.	810	390	190	110	110	60	60	60	60	90	1410
.50	a.	600	330	200	140	100	100	60	50	50	50	40
(1.15)	b.	540	310	190	140	110	110	80	80	100	110	130
.75	a.	1590	700	290	230	170	130	90	100	90	70	80
(1.73)	b.	720	290	180	130	110	130	90	90	140	230	240
.90	a.	2670	1110	750	440	290	180	150	110	80	70	70
(2.07)	b.	290	200	150	120	100	60	50	50	120	100	160

(b) n=100

p (t_p) \ h		.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55
.10	a.	500	190	120	200	340	550	740	960	1130	1300	1450
(.23)	b.	510	200	120	220	400	640	880	1150	1350	1560	1750
.25	a.	380	140	90	60	50	30	30	40	60	90	130
(.58)	b.	380	140	90	60	60	30	30	30	50	80	130
.50	a.	350	160	90	60	50	40	30	30	30	30	30
(1.15)	b.	340	160	90	70	60	50	50	70	100	110	110
.75	a.	530	220	120	110	90	70	70	70	60	60	60
(1.73)	b.	340	140	90	80	80	70	100	80	90	150	250
.90	a.	2130	950	390	310	200	150	110	90	70	60	50
(2.07)	b.	420	150	110	80	80	70	110	140	240	260	380

a. MSE $\hat{f}_n$, b. MSE f_n^*

Table 4.9. Estimated MSE of Kernel Density Estimators

Life Distribution: W(2,1), Censoring Distribution: $U(0,t_{.90})$
Kernel: N(0,1)
(All entries are to be multiplied by 1.0E-04.)

(a) n=50

p (t_p)	h	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55
.10	a.	390	150	100	80	80	60	70	70	60	60	60
(.15)	b.	380	140	100	70	50	50	40	40	40	40	50
.25	a.	850	460	230	180	170	210	210	240	280	320	370
(.38)	b.	850	460	260	250	290	430	480	600	740	870	980
.50	a.	1880	810	400	240	240	220	290	350	460	600	770
(.76)	b.	1770	830	400	290	290	350	450	580	780	920	1130
.75	a.	3020	1230	580	290	170	120	50	40	30	60	100
(1.14)	b.	2150	770	680	440	350	750	640	620	470	850	480
.90	a.	6280	1860	910	450	280	170	110	80	50	30	20
(1.37)	b.	1890	910	590	530	650	90	1250	1920	2550	3380	3360

(b) n=100

p (t_p)	h	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55
.10	a.	180	80	50	50	40	40	40	40	40	50	40
(.15)	b.	180	80	50	50	30	30	20	30	30	30	40
.25	a.	510	230	130	100	120	150	180	230	270	330	380
(.38)	b.	500	240	140	160	250	350	460	590	720	850	980
.50	a.	870	340	170	150	140	170	250	340	470	620	790
(.76)	b.	840	330	180	170	170	230	360	520	700	900	1080
.75	a.	1100	620	340	160	110	60	30	20	30	50	90
(1.14)	b.	1020	510	400	260	300	400	390	420	420	350	330
.90	a.	3640	1410	520	320	190	110	80	50	30	20	10
(1.37)	b.	1350	610	640	840	1450	2760	3540	4720	7980	5760	7700

a. MSE $\hat{f}_n$, b. MSE f_n^*

Table 4.10. Estimated MSE of Kernel Density Estimators

Life Distribution: W(.5,1), Censoring Distribution: $U(0,t_{.75})$
Kernel: N(0,1)
(All entries are to be multiplied by 1.0E-04.)

(a) n=50

p (t_p) \ h		.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55
.10	a.	880	550	410	170	100	160	300	470	660	800	990
(.19)	b.	900	640	500	200	120	160	290	490	690	860	1080
.25	a.	660	210	160	160	170	170	160	120	90	80	50
(.48)	b.	640	210	170	200	260	320	300	260	180	160	100
.50	a.	440	200	150	130	90	60	70	110	120	170	190
(.96)	b.	400	190	140	130	90	70	110	140	190	280	330
.75	a.	1540	530	520	430	390	360	330	330	310	270	270
(1.44)	b.	600	220	160	130	110	70	60	100	130	240	300
.90	a.	7440	5040	2700	1930	1190	910	700	540	450	330	300
(1.73)	b.	470	270	180	130	130	90	100	150	200	310	330

(b) n=100

p (t_p) \ h		.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55
.10	a.	420	370	260	120	60	120	270	440	640	820	980
(.19)	b.	420	450	340	160	60	110	270	450	660	890	1070
.25	a.	200	130	80	90	130	150	140	100	70	50	40
(.48)	b.	200	130	90	130	220	290	270	230	180	130	80
.50	a.	210	100	50	40	40	40	50	70	90	130	160
(.96)	b.	200	100	50	50	50	60	90	140	180	250	320
.75	a.	750	190	150	140	190	200	250	240	240	250	240
(1.44)	b.	360	140	70	70	40	50	60	70	80	160	260
.90	a.	4320	2410	2090	1530	1170	820	650	490	410	340	280
(1.73)	b.	330	170	170	100	130	110	170	180	360	470	650

a. MSE $\hat{f}_n$, b. MSE f_n^*

Table 4.11. Estimated MSE of Kernel Density Estimators

Life Distribution: W(.5,1), Censoring Distribution: E(1)
Kernel: N(0,1)
(All entries are to be multiplied by 1.0E-04.)

(a) n=50

p (t_p) \ h		.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55
.10	a.	1390	650	400	900	1670	2370	3090	3540	4040	4580	4930
(.11)	b.	1520	780	390	920	1730	2560	3300	3850	4410	5000	5470
.25	a.	820	380	400	350	200	110	60	50	70	130	180
(.29)	b.	820	420	590	590	410	220	140	70	60	90	150
.50	a.	550	190	150	130	110	120	90	130	140	120	140
(.69)	b.	510	180	150	150	150	310	290	450	590	550	660
.75	a.	680	520	240	210	130	120	110	110	90	90	100
(1.39)	b.	530	280	150	120	80	60	90	90	110	140	270
.90	a.	1950	650	510	370	350	200	180	150	120	120	100
(2.30)	b.	420	150	90	70	80	50	50	30	40	40	30

(b) n=100

p (t_p) \ h		.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55
.10	a.	980	360	270	830	1670	2350	2910	3630	4040	4480	4870
(.11)	b.	1090	440	260	830	1730	2520	3120	3960	4400	4900	5330
.25	a.	310	260	260	230	140	80	30	30	50	100	150
(.29)	b.	310	310	430	470	330	190	100	40	30	60	110
.50	a.	310	120	90	50	60	80	80	110	110	100	90
(.69)	b.	300	120	90	60	130	230	310	470	490	510	520
.75	a.	380	270	100	70	60	50	40	50	40	40	40
(1.39)	b.	320	160	80	60	50	50	50	60	90	150	220
.90	a.	670	420	290	230	140	120	100	90	60	50	50
(2.30)	b.	200	100	80	60	50	40	40	30	30	20	20

a. MSE $\hat{f}_n$, b. MSE f^*_n

Table 4.12. Comparison of $\hat{f}_n$ and f_n^* for Small Samples

K - Standard Normal $\qquad h(n) = n^{-1/5}$

Life Distribution: E(5) $\qquad$ Censoring Distribution: E(1)

(All entries to be multiplied by 10^{-4}.) $\qquad a = \frac{1}{6}$

$f^o(t)$ (t)	n	$\hat{f}_n(t)$ Bias	$\hat{f}_n(t)$ Variance	$\hat{f}_n(t)$ MSE	$f_n^*(t)$ Bias	$f_n^*(t)$ Variance	$f_n^*(t)$ MSE
1900 (2600)	10	- 279	111	119	- 848	46	118
	20	- 639	449	856	- 912	31	114
	50	- 625	223	619	- 815	18	85
	100	- 552	136	440	- 707	12	62
1500 (14400)	10	1552	294	534	175	129	132
	20	816	288	362	195	127	131
	50	84	83	83	129	65	66
	100	- 2	37	37	116	37	39
1000 (34700)	10	242	230	236	- 778	35	96
	20	770	314	373	- 714	44	95
	50	1435	406	612	- 480	81	104
	100	1259	410	568	- 339	86	98
500 (69300)	10	- 474	6	28	- 497	1	25
	20	- 464	9	30	- 498	0	25
	50	- 385	48	63	- 496	1	25
	100	- 373	41	55	- 489	2	26

Table 4.13. Comparison of $\hat{f}_n$ and f_n^* for Small Samples

K - Standard Normal $\quad h(n) = n^{-1/5}$

Life Distribution: E(1) $\quad$ Censoring Distribution: E(10)

(All entries to be multiplied by 10^{-4}.) $\quad a = \frac{10}{11}$

$f^o(t)$ (t)	n	$\hat{f}_n(t)$ Bias	$\hat{f}_n(t)$ Variance	$\hat{f}_n(t)$ MSE	$f_n^*(t)$ Bias	$f_n^*(t)$ Variance	$f_n^*(t)$ MSE
9500 (513)	10	- 6127	65	3819	- 6242	65	3961
	20	- 5913	41	3538	- 6022	41	3668
	50	- 5671	24	3224	- 5771	24	3354
	100	- 5463	17	3001	- 5554	17	3101
7500 (2900)	10	- 3568	66	1339	- 3654	68	1402
	20	- 3222	42	1080	- 3297	44	1131
	50	- 2779	24	797	- 2832	25	827
	100	- 2419	18	602	- 2458	18	622
5000 (6931)	10	- 865	47	122	- 873	52	128
	20	- 562	32	63	- 541	36	65
	50	- 241	19	25	- 205	19	24
	100	- 73	13	14	- 35	14	14
2500 (1390)	10	383	46	61	408	48	64
	20	359	32	45	404	34	51
	50	274	18	25	325	19	29
	100	203	11	16	240	12	17
500 (29957)	10	107	26	27	56	22	22
	20	60	15	15	29	14	14
	50	59	8	8	49	8	8
	100	37	4	4	36	4	5

Table 4.14. Comparison of $\hat{f}_n$ and f_n^* for Small Samples

K - Uniform [-1,1] $h(n) = n^{-1/5}$

Life Distribution: E(5) Censoring Distribution: E(1)

(All entries to be multiplied by 10^{-4}.) $a = \frac{1}{6}$

$f^o(t)$ (t)	n	$\hat{f}_n(t)$ Bias	Variance	MSE	$f_n^*(t)$ Bias	Variance	MSE
1900 (2600)	10	- 322	176	186	- 622	86	124
	50	- 480	42	65	- 582	34	68
	100	- 381	25	39	- 450	21	42
	300	- 186	11	14	- 208	10	14
1500 (14400)	10	1773	840	1153	- 103	289	290
	50	- 16	148	148	- 53	122	122
	100	- 4	72	72	1	70	70
	300	- 10	32	32	4	32	32
1000 (34700)	10	142	526	528	- 808	65	130
	50	1543	1060	1297	- 1044	144	173
	100	1320	976	1149	- 415	159	176
	300	221	324	329	- 123	145	147
500 (69300)	10	- 479	9	32	- 495	1	26
	50	- 376	94	108	- 500	0	25
	100	- 398	69	84	- 486	5	29
	300	- 241	179	185	- 466	13	35

Table 4.15. Comparison of $\hat{f}_n$ and f_n^* for Small Samples

K - Uniform [-1,1] $h(n) = n^{-1/5}$

Life Distribution: E(1) Censoring Distribution: E(10)

(All entries to be multiplied by 10^{-4}.) $a = \frac{10}{11}$

$f^o(t)$ (t)	n	$\hat{f}_n(t)$ Bias	$\hat{f}_n(t)$ Variance	$\hat{f}_n(t)$ MSE	$f_n^*(t)$ Bias	$f_n^*(t)$ Variance	$f_n^*(t)$ MSE
9500 (513)	10	- 5570	164	3266	- 5667	158	336[illegible]
	50	- 5173	61	2736	- 5254	58	2819
	100	- 4987	38	2526	- 5057	37	2595
	300	- 4640	18	2171	- 4699	18	2225
5000 (6931)	10	332	152	163	361	160	173
	50	210	62	66	237	63	69
	100	146	42	44	166	43	46
	300	70	18	19	86	19	19
2500 (1390)	10	174	158	161	114	159	160
	50	100	45	46	108	46	47
	100	56	26	26	62	26	27
	300	35	13	13	42	13	13
500 (29957)	10	25	52	52	- 31	43	43
	50	21	15	15	2	14	14
	100	28	8	8	- 6	8	8
	300	8	3	3	7	3	3

standard normal density seems to be the best choice. Other kernel functions which closely fit the standard normal (Parzen, 1962) may perform as well but were not included in this study. The estimator $\hat{f}_n$ is fairly robust with respect to the life distribution, and $\hat{f}_n$ performs well near the "center" of the life distribution regardless of whether the censoring distribution is exponential or uniform.

From the results represented by Tables 4.8-4.11, it is evident that for each x, the estimated MSE appears to have at least a relative minimum at some value of h. These seem to occur near the values $n^{-1/2}$, $n^{-1/3}$, or $n^{-1/5}$, although it seems to be difficult to prove this result analytically, as mentioned before. Based on these results and the results represented by Tables 4.1-4.7 that the estimated variances of $\hat{f}_n$ increase as h ranges from $n^{-1/5}$ to $n^{-1/2}$, the value $h = n^{-1/5}$ or $n^{-1/3}$ seems to be a reasonable choice for the bandwidth in practice.

The above conclusions indicate that the bias, variance, and mean squared error of $\hat{f}_n(t)$ decrease as n becomes larger, regardless of the life distribution or censoring distribution. This leads to the conjecture that the asymptotic results of McNichols and Padgett (1981) hold <u>without</u> the condition of the Koziol-Green (or proportional hazards) model of random censorship. Most of the cases simulated do not satisfy the condition of that model. An analytical proof of this conjecture, however, would be quite difficult.

The simulation results represented by Tables 4.8-4.11 indicate that, with respect to estimated mean squared error, f_n^* and $\hat{f}_n$ behave similarly as the bandwidth values vary. The simulations (Tables 4.12-4.15) also indicate that the Blum-Susarla estimator f_n^* and the estimator $\hat{f}_n$ perform about the same with respect to bias, variance, and mean squared error when a = P(uncensored observation) is larger than 0.5. When $a < 1/2$, f_n^* tends to have smaller variance and mean squared error than $\hat{f}_n$. This can probably be explained by noting that f_n^* depends upon having a good estimate of the censoring survival function $1-H$ in the denominator, and when there is a large portion of the observations which are censored, the denominator H_n^* of (3.1) would give a good estimate of $1-H$. Hence, when a is small, f_n^* would generally provide a better density estimate than $\hat{f}_n$ with respect to smaller variance and mean squared error.

ACKNOWLEDGMENTS

The authors are grateful to the referee and Associate Editor for their very helpful suggestions which improved this paper and to Clara Day for help in computer programming for the results of Section 4. The first author was supported during the revision of this paper by the U.S. Air Force Office of Scientific Research and Army Research Office under grant number AFOSR-84-0156.

REFERENCES

Bean, S. J. and Tsokos, C. P. (1980). Developments in nonparametric density estimation. Intern. Statist. Rev. 48, 215-235.

Blum, J. R. and Susarla, V. (1980). Maximal deviation theory of density and failure rate function estimates based on censored data. Multivariate Analysis - V, P. R. Krishniah, Ed., North-Holland, New York, 213-222.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. Ann. Statist. 2, 437-453.

Chen, Y. Y., Hollander, M., and Langberg, N. A. (1982). Small-sample results for the Kaplan-Meier estimator. J. Amer. Statist. Assoc. 77, 141-144.

Cox, D. R. (1972). Regression models and life-tables. J. Royal Statist. Soc. B34, 187-220.

Csörgő, S. and Horváth, L. (1981). On the Koziol-Green model for random censorship. Biometrika 68, 391-401.

Efron, B. (1967). The two sample problem with censored data. Proc. Fifth Berkeley Symp. Math. Statist. Prob. 4, 831-853.

Földes, A. Rejtö, L., and Winter, B. B. (1980). Strong consistency properties of nonparametric estimators for randomly censored data, I: The product-limit estimator. Periodica Mathematica Hungarica 11, 233-250.

Földes, A., Rejtö, L., and Winter, B. B. (1981). Strong consistency properties of nonparametric estimators for randomly censored data. Part II: Estimation of density and failure rate. Periodica Mathematica Hungarica 12, 15-29.

Fryer, M. J. (1977). A review of some non-parametric methods of density estimation. J. Inst. Maths. Applics. 20, 335-354.

International Mathematical and Statistical Libraries (1980). IMSL, Inc., 7500 Bellaire Blvd., Houston, Texas, USA.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53, 457-481.

Koziol, J. A. and Green, S. B. (1976). A Cramer-von Mises statistic for randomly censored data. Biometrika 63, 465-473.

Lagakos, S. W. (1979). General right censoring and its impact on the analysis of survival data. Biometrics 35, 139-156.

McNichols, D. T. and Padgett, W. J. (1981). Kernel density estimation under random censorship. Statistics Tech. Rep. No. 74, University of South Carolina.

McNichols, D. T. and Padgett, W. J. (1984). A modified kernel density estimator for randomly right-censored data. S. African Statist. J. 18, 13-27.

Michael, J. R., Schucany, W. R. and Haas, R. W. (1976). Generating random variates using transformations with multiple roots. Amer. Statist. 30, 88-90.

Miller, R. G. (1981). Survival Analysis. John Wiley & Sons, New York.

Padgett, W. J. and McNichols, D. T. (1984). Nonparametric density estimation from censored data. Commun. Statist.-Theor. Meth. B, 1581-1611.

Parzen, E. (1962). On estimation of a probability density function and mode. Ann. Math. Statist. 33, 1065-1076.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. Ann. Math. Statist. 27, 832-837.

Rosenblatt, M. (1976). On the maximal deviation of k-dimensional density estimates. Ann. Prob. 4, 1009-1015.

Scott, D. W. and Factor, L. E. (1981). Monte Carlo study of three data-based nonparametric probability density estimators. J. Amer. Statist. Assoc. 76, 9-15.

Tapia, R. A. and Thompson, J. R. (1978). Nonparametric Probability Density Estimation. The Johns Hopkins Univ. Press, Baltimore.

Wegman, E. J. (1972a). Nonparametric probability density estimation, I. A summary of available methods. Technometrics 14, 533-546.

Wegman, E. J. (1972b). Nonparametric density estimation: II. A comparison of density estimation methods. J. Statist. Comput. Simul. 1, 225-245.

Wellner, J. (1982). Asymptotic optimality of the product limit estimator. Ann. Statist. 10, 595-602.

Wertz, W. and Schneider, B. (1979). Statistical density estimation: a bibliography. Intern. Statist. Rev. 47, 155-175.

END

FILMED

1-85

DTIC